**ANTICIPATING BUSINESS BANKRUPTCY**

The field of bankruptcy prediction, a prominent topic in economics for almost a century, focuses on creating predictive models that use various econometric indicators to forecast a company's financial health. These models assist in evaluating a company's financial status and its long-term prospects in the market. The research discussed here involves analysing businesses in Poland from 2000 to 2012, with a specific emphasis on those that went bankrupt and those that remained operational in 2007. The dataset includes financial ratios from the initial year, with class labels indicating bankruptcy status. These labels are used to categorize companies based on their vulnerability to bankruptcy.

The bankruptcy prediction field has evolved to keep pace with the changing global economic landscape, employing increasingly sophisticated tools and methodologies. Researchers continuously improve their approaches to provide businesses and stakeholders with critical insights into their financial health and long-term viability in the market. This area of study remains vital for decision-makers seeking to anticipate and manage financial risks in the corporate world.

**Aim:**

Develop a predictive model for anticipating business bankruptcy using econometric measures and historic data.

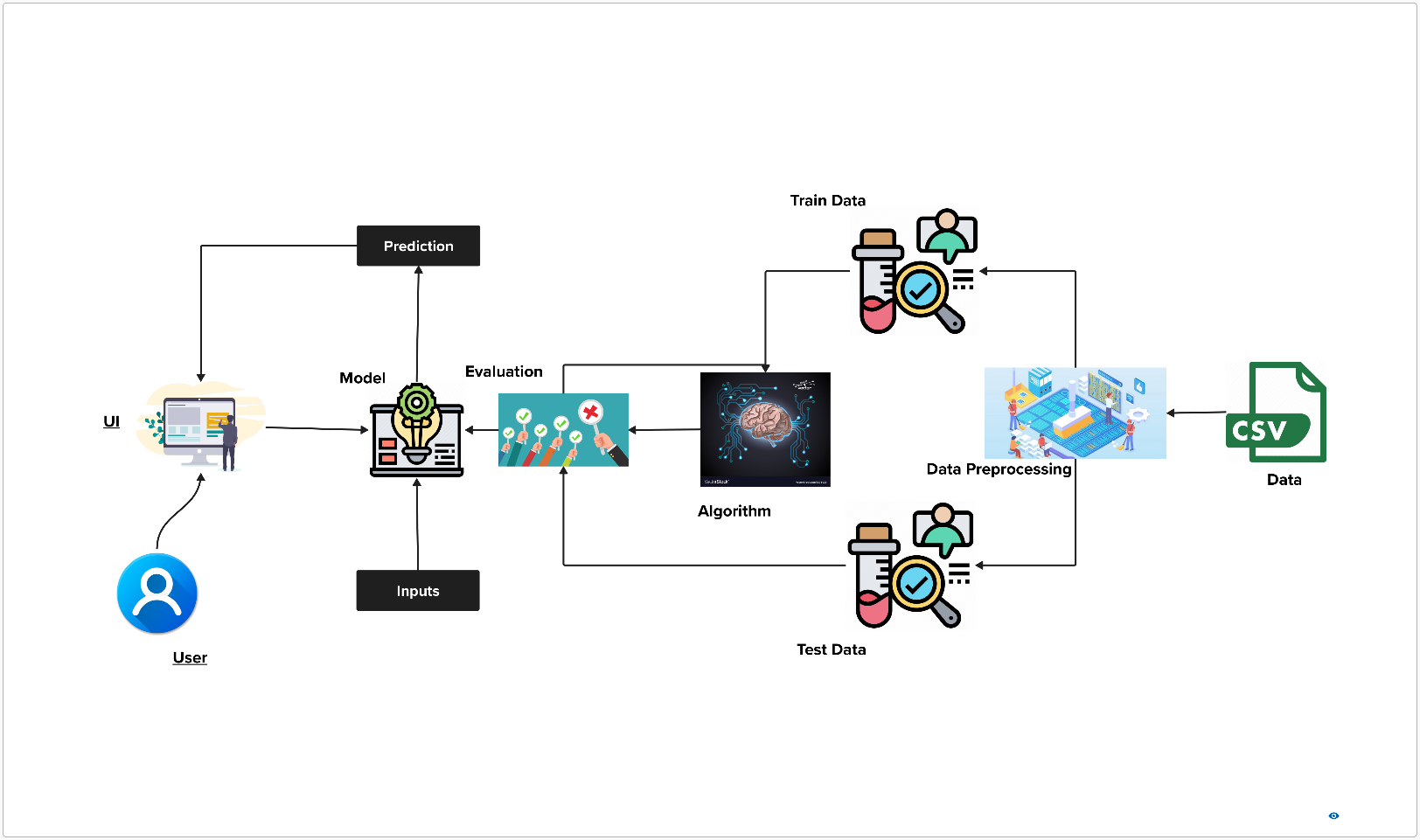
**INDEX**

|  |
| --- |
| 1. [TECHNICAL ARCHITECTURE](#tech_arch) |
| 1. [PROJECT FLOW](#PROJECT_FLOW) |
| 1. [PRIOR KNOWLEDGE](#prior_knowledge) |
| 1. [PROJECT STRUCTURE](#PROJECT_STRUCTURE) |
| 1. [DEFINING THE PROBLEM](#DEFINING_THE_PROBLEM)    1. Objective    2. Business Requirements    3. Business Impact    4. Social Impact |
| 1. [DATA COLLECTION & PREPARATION](#DATA_COLLECTION_AND_PREPARATION)    1. Collecting the dataset    2. Data Preparation    3. Data Preprocessing 2. Exploratory Data Analysis |
| 1. [MODEL BUILDING](#MODEL_BUILDING)    1. Decision Tree Classifier    2. Random Forest Classifier    3. XGBoost Classifier |
| 1. PERFORMANCE TESTING |
| 1. MODEL DEPLOYMENT    1. Testing the Model    2. Webpage Creation & Flask Deployment |
| 10. PROJECT DEMONSTRATION   1. Implementation 2. Video Explanation |

**TECHNICAL ARCHITECTURE**

In the process of building a machine learning model, several key steps are followed. It begins with the raw CSV data, which serves as the foundational dataset. Data preprocessing is a critical phase involving data cleaning to address missing values, outliers, and inconsistencies, as well as data transformation, which encompasses feature engineering, encoding categorical variables, scaling, and normalization. Feature selection is performed to choose relevant attributes. The dataset is then split into training data (80% of the dataset) for model training and test data (20%) for model evaluation.

Algorithm selection depends on the task type and data characteristics, with the chosen algorithm used to train the model. In our project, we deal with a classification problem and hence, will be using Classification algorithms in ML such as, Decision Trees, Random Forest, Extreme Gradient Boosting, etc. Model performance is assessed using metrics like accuracy scores and classification reports. A decision point is reached after evaluation: if the model's performance is unsatisfactory, it returns to algorithm selection and data preparation; if it meets criteria, the final model is built using the entire dataset. Successful models can be deployed in the real world, with continuous monitoring and adjustments to ensure ongoing performance as new data becomes available. This comprehensive process ensures the creation of effective machine learning models for various tasks.



**Figure 1.1 The diagram of the technical architecture: Image**

**PROJECT FLOW**

To ensure the successful execution of our project, it is crucial that we comprehensively outline the major flow of activities, tasks, and processes involved. The entire project can be broken down into a sequence of steps, each contributing to the overall goal of developing an interactive and intelligent user interface driven by predictive and analytical models. Here's an expanded overview of the project's major flow and the associated tasks:

1. Project Initiation:

* Define the project scope and objectives.
* Form the project team and allocate responsibilities.
* Establish a clear timeline and milestones for project execution.

1. Problem Definition:

* Begin with a thorough understanding of the problem statement.
* Specify the business problem to be addressed through the UI and integrated model.
* Identify and document the specific business requirements that the UI should meet.
* Highlight the potential social and business impact of solving the problem through the project, addressing its broader implications.

1. Data Collection:

* Identify sources of relevant data for the project.
* Develop a data collection strategy and gather the necessary datasets.
* Ensure proper storage and organization of the collected data for later use in the project.

1. Data Preparation:

* Preprocess the collected data, which includes:
* Handling null or missing values, employing techniques like imputation or removal.
* Extracting relevant features from the raw data, preparing it for analysis.

1. Exploratory Data Analysis:

* Conduct data visualization and exploratory data analysis to gain insights into the dataset.
* Visualize data distributions, correlations, and patterns to inform subsequent modelling decisions.

1. Data Segmentation (X and Y Splitting):

* Split the dataset into input (X) and target (Y) variables, enabling the model to learn from the data.

1. Feature Selection:

* Identify and select the most relevant features to be used in the model, enhancing its efficiency and effectiveness.

1. Train and Test Data Splitting:

* Split the dataset into training and testing subsets to assess the model's performance accurately.

1. Data Balancing:

* Address any class imbalance issues in the dataset to ensure that the model doesn't favour one class over others.

1. Model Building:

* Develop the predictive and analytical model using appropriate algorithms and techniques.
* Train the model on the training data to learn patterns and relationships.

1. Performance Testing:

* Assess the model's performance using various evaluation metrics, such as:
* Accuracy scores to measure overall model correctness.
* Classification reports to provide insights into precision, recall, and F1-score.

1. Model Saving and Export:

* Save the trained model for future use and export it for deployment.

1. Deployment of the Model:

* Develop a user-friendly GUI for interaction with the model.
* Integrate the GUI with a web framework, allowing users to access the UI through web browsers.
* Deploy the project, making it accessible to the intended audience, and ensure it runs smoothly in a production environment.

Addressing each of these project activities and tasks comprehensively, will enable us to create a seamless and intelligent user interface that leverages predictive and analytical models to address the specified business problem effectively. This systematic approach will significantly enhance our chances of a successful project implementation.

**PRIOR KNOWLEDGE**

To make sure the bankruptcy prediction project works well, it's really important for the project team to know a lot about five important things. These things are crucial for the project:

1. *Financial Domain Proficiency*:

* Know a lot about financial numbers, like balance sheets and income statements.
* This knowledge helps understand how well a company is doing financially.
* Be familiar with financial terms like how quickly a company can get cash, how much money it's making, if it can pay its debts, and how efficiently it uses its assets.

1. *Machine Learning Foundations and Classification Algorithms*:

* It's crucial to know the basics of machine learning tools like decision trees, random forests, and extreme gradient boosting classification algorithms.
* Knowing the fine details of these tools, like making them work even better and using a bunch of them together, is a big help in making the models work well and give the right answers.

1. *Feature Selection Expertise*:

* Requires effective selection of relevant features, with a specific focus on Recursive Feature Elimination (RFE).
* Choosing the right stuff from the data makes it simpler, makes the model work better, and makes training the model faster.
* Hence, it's very important to be good at finding and keeping the most useful information.

1. *Imbalanced Data Handling Skills*:

* Dealing with imbalanced data is a will be a crucial step.
* It's important to be familiar with techniques like Synthetic Minority Over-sampling Technique (SMOTE) to fix this problem.
* Being unable to handle this issue, will result in incorrect predictions.

1. *Data Source Awareness*:

* Understanding the data source is vital for the project's integrity.
* In this context, data was collected from Emerging Markets Information Service (EMIS), a comprehensive database of emerging markets worldwide.
* Being well-informed about the data source provides the foundation for conducting a reliable analysis.

By honing their expertise in these key domains, the project team can effectively pave the way for the development of a robust bankruptcy prediction model. With this multifaceted knowledge, they can navigate the complexities of financial data, machine learning techniques, feature selection, data balancing, and data source intricacies to create a predictive model that offers valuable insights into a company's financial stability.

**PROJECT STRUCTURE**

**DEFINING THE PROBLEM**

**OBJECTIVE**

The objective of bankruptcy prediction is to assess both a company's current financial condition and its future outlook within the framework of its continued presence and operations in the market over the long term.

**BUSINESS REQUIREMENTS**

* *Risk Assessment*: Through the anticipation of bankruptcy likelihood, businesses gain the ability to assess the financial hazards linked to engaging with a specific company. This equips them to make knowledgeable choices regarding extending credit, entering contractual agreements, or venturing into collaborative endeavors.
* *Financial Strategy*: Precise bankruptcy prediction plays a pivotal role in a company's financial strategy. It empowers organizations to foresee possible challenges in cash flow, identify areas in need of enhancement, and proactively institute measures to diminish vulnerabilities.
* *Early Detection Mechanism*: Models for forecasting bankruptcy can act as an early detection mechanism, providing companies with timely alerts about potential financial troubles before they escalate. This enables companies to promptly implement corrective actions, including debt restructuring, contract renegotiation, or the implementation of cost-cutting initiatives.

**BUSINESS IMPACT**

* *Credit Risk Management*: The ability to accurately predict bankruptcy helps banks and financial institutions assess creditworthiness, make informed lending decisions, and manage credit risk, reducing the likelihood of loan defaults and financial losses.
* *Investment Decisions*: Investors can use bankruptcy predictions to evaluate the financial health of companies before making investments, which can include stock purchases, bond investments, and decisions related to initial public offerings (IPOs).
* *Supply Chain Management*: Proactive identification of financially distressed suppliers allows companies to secure alternative sources and maintain a smooth and resilient supply chain.
* *Mergers and Acquisitions*: Companies can assess the financial risk associated with potential merger or acquisition targets, aiding in deal negotiations and structuring transactions.
* *Regulatory Compliance*: Businesses in regulated industries can use bankruptcy prediction to demonstrate compliance with financial stability requirements imposed by regulators, ensuring adherence to industry standards and avoiding regulatory penalties.

**SOCIAL IMPACT**

* *Job Security***:** When businesses can predict financial problems early, they're more likely to take steps to prevent bankruptcy. This helps protect the jobs of employees and ensures a more stable work environment.
* *Customer Trust***:** Maintaining financial stability and a low risk of bankruptcy can positively impact a company's reputation and customer trust. Customers feel more confident doing business with companies that are financially secure.
* *Economic Stability***:** By reducing the likelihood of businesses going bankrupt, the project contributes to overall economic stability. Fewer bankruptcies mean fewer disruptions in the job market and a healthier economy.
* *Supplier Relationships***:** The project helps companies maintain good relationships with suppliers by ensuring they don't run into financial trouble. This, in turn, fosters stronger social and business connections.
* *Community Well-being***:** Strong, stable businesses contribute to the well-being of the communities in which they operate. They can support local causes, provide employment, and stimulate economic growth, benefiting the overall quality of life in the community.

**DATA COLLECTION AND PREPARATION**

**COLLECTING THE DATASET**

The dataset utilized in the project was sourced from Kaggle, a prominent platform for data sharing and analysis. This dataset consisted of five distinct CSV files, each corresponding to a specific year within the time frame from 2007 to 2012. These files collectively provided the foundational data required for the bankruptcy prediction project mentioned earlier.

Kaggle dataset files: [Link](https://www.kaggle.com/datasets/bhadaneeraj/bankruptcy-detection)

Each of the five dataset files, spanning from 2007 to 2012, contains 65 columns. The critical column, "class," serves as the target variable for predicting a company's financial status. The remaining 64 columns, labeled "Attr1" through "Attr64," encompass diverse financial metrics and ratios, providing a comprehensive financial overview. This dataset forms the foundation for our bankruptcy prediction project, enabling in-depth analysis and predictive modeling.

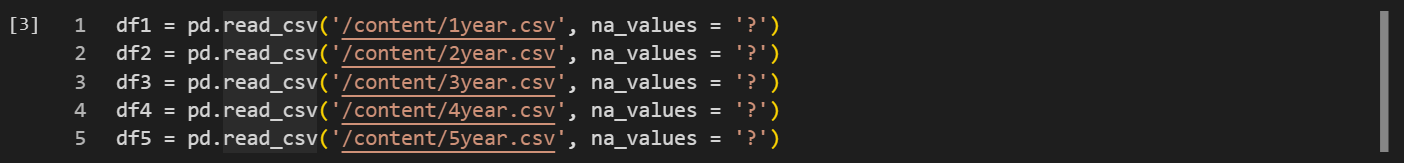
**DATA PREPARATION**

1. *Importing the libraries*

*A screenshot of a computer program

Description automatically generated*

1. *Read the dataset*

**

*A black and grey rectangle

Description automatically generated*

1. *Combining the files into one data frame*

**DATA PREPROCESSING**

1. *Checking column information*
2. *Dropping columns with large numbers of missing values*
3. *Replacing the missing/null values*

**EXPLORATORY DATA ANALYSIS**

1. *Combin*
2. *Vwegwq*
3. *Werger*

**MODEL BUILDING**

**PERFORMANCE TESTING**

**MODEL DEPLOYMENT**

1. *Decision Tree Classifier:*

* bvrtb

1. *Random Forest Classifier:*

* s

1. *Extreme Gradient Boosting Classifier:*

* dsfb

**PROJECT DEMONSTRATION**